УДК 004.8:378.147
DOI: https://doi.org/10.64076/ihrc250630.03

**Melnyk, H.,**
PhD in Economics, Associate
Professor of the Department
of Applied Mathematics and IT,
Yuriy Fedkovych Chernivtsi
National University, Chernitsi
**Melnyk, V.,**
PhD in Physics and Mathematics,
Assistant of the Department
of Mathematical Modeling
Yuriy Fedkovych Chernivtsi
National University, Chernitsi



# A HYBRID FUZZY-GENERATIVE FRAMEWORK FOR AUTOMATED ASSESSMENT OF IMPLICIT STUDENT COMPETENCIES USING MOODLE CHAT LOGS

In today's rapidly evolving higher-education landscape, the ability to assess and nurture students' *soft competencies* — such as critical thinking, effective communication, and emotional intelligence — has become as crucial as measuring their mastery of domain-specific knowledge. Traditional assessment methods (multiple-choice tests, static assignments) excel at evaluating factual recall and procedural skills, but they fall short of capturing these nuanced, non-cognitive attributes. At the same time, modern learning management systems (LMS) generate rich streams of student–instructor and peer-to-peer interactions in discussion forums and real-time chats. These text-based artifacts contain latent signals about learners' reasoning patterns, argumentation strength, and collaborative behaviors.

Recent advances in Natural Language Processing (NLP), in particular transformer-based sentence embeddings (e.g., BERT, RoBERTa) and sentiment-analysis networks, offer powerful means to extract semantic and affective features from free-form text. Concurrently, fuzzy logic provides a principled framework for modeling human-centric, linguistically expressed concepts — such as "high argumentation" or "moderate engagement" — that naturally resist crisp numerical thresholds. Moreover, generative neural architectures (variational autoencoders, generative adversarial networks) enable data augmentation in low-resource scenarios, synthesizing diverse "virtual" student responses to bolster model robustness.

This paper proposes a *hybrid fuzzy-generative framework* for the automated evaluation of students' implicit competencies, leveraging:

1. *NLP feature extraction*, including sentiment probabilities and topic-model distributions from open responses and chat logs;
2. *Fuzzy inference rules*, which map numerical indicators (e.g., sentiment scores, lexical diversity) into intuitive linguistic categories ("low", "medium", "high");
3. *Generative modeling*, which augments scarce labeled examples by producing synthetic embeddings representing various competency levels.

By integrating these components, our approach aims to produce more flexible, transparent, and context-aware assessments of soft skills directly within the LMS environment. We validate the framework on real-world Moodle chat-log data collected via an university's portal, comparing its outputs against expert human ratings. Results demonstrate improved agreement (Cohen's $\kappa > 0.75$) and classification accuracy (+12 pp) compared to baseline methods lacking either the fuzzy or generative component.

Natural Language Processing (NLP) has been extensively applied to analyze student-generated text in educational contexts. Bird's foundational work with the NLTK toolkit demonstrates core techniques — tokenization, part-of-speech tagging, and parsing — that underpin modern text analytics pipelines for learner data [1]. More recently, Shaik surveys advancements and persistent challenges in adopting transformer-based and neural architectures (e.g., BERT, GPT) for automated feedback analysis, highlighting issues of data sparsity and domain adaptation in educational settings [3].

Accurate extraction of aspect terms like "clarity" or "argument strength" is essential for competency inference. Gandhi et al. show that a hybrid CRF + bi-LSTM model outperforms standalone methods in this task, informing our feature-engineering pipeline [4], while Richardson's review highlights the value of combining quantitative surveys with open-text prompts [2]. Although robust NLP workflows and feedback instruments exist [1, 4, 2], they typically use hard thresholds or classifiers; Shaik calls for more flexible, interpretable mechanisms [3], and to our knowledge no prior work unites fuzzy inference with generative augmentation for implicit competency modeling. Our approach addresses this gap by uniting state-of-the-art NLP extraction, fuzzy-rule interpretation, and synthetic data generation within a single framework.

A research protocol was developed to create and evaluate a hybrid fuzzy-generative framework for assessing implicit competencies through analysis of student chat logs (Figure 1).


**Figure 1.** Research protocol flowchart for the hybrid fuzzy-generative framework for implicit competency assessment using Moodle chat logs.

**I. Data Collection and Preprocessing**

• *Source Platform:* We export anonymized chat and forum logs from Moodle, capturing student ID, role, timestamp, and raw message text.

• *Preprocessing Pipeline:*

1. Cleaning: strip HTML tags and special characters, remove stop-words via NLTK routines [1].
2. Normalization: lowercase, lemmatize with WordNet, and segment sentences.
3. Deduplication: eliminate duplicate posts and system notifications.

**II. NLP Feature Extraction**

1. Semantic Embeddings: fine-tuned BERT generates 768-dimensional embeddings representing message semantics.

2. Aspect Extraction: a bi-LSTM + CRF tagger [4] identifies key terms tied to soft competencies (e.g., “argumentation,” “collaboration”), recording frequency and confidence.

3. Sentiment & Engagement: a bidirectional LSTM classifier outputs positive/neutral/negative sentiment scores; we also compute message length, response latency, and reply density.

The resulting feature vector for each student session concatenates embeddings, aspect metrics, sentiment probabilities, and engagement indicators.

**III. Fuzzy Inference Design**

• *Membership Functions:* define linguistic variables (e.g., Sentiment_Positive ∈ {Low, Medium, High}) with triangular functions over normalized ranges.

• *Rule Base:* expert-crafted IF–THEN rules combine variables, for example:

IF (Sentiment_Positive IS High) AND (Argument_Frequency IS Frequent)
THEN Critical_Thinking IS High

• *Defuzzification:* use the centroid method to convert aggregated fuzzy outputs into crisp scores (0–100) for each competency dimension.

**IV. Generative Model Training**

• *Conditional VAE Architecture:*

○ Encoder: maps the concatenated semantic embeddings and auxiliary features to a latent Gaussian distribution.

○ Latent Sampling & Decoder: reconstructs the original feature vector, enabling synthetic example generation.

• *Loss Components:* combine reconstruction loss (MSE) with KL divergence regularization.

• *Data Augmentation:* generate synthetic feature vectors for underrepresented competency levels to enhance training diversity and model robustness.

To assess the efficacy of the proposed hybrid fuzzy–generative framework, we conducted experiments on a dataset comprising anonymized chat logs from two undergraduate courses hosted in Moodle. A total of 1 200 discussion threads and 4 500 individual messages were sampled, each annotated by three expert raters according to predefined soft-skill levels (low, medium, high) for critical thinking and collaboration. Inter-rater agreement exceeded 0.80 (Fleiss' $\kappa$), ensuring reliable ground truth for subsequent comparison.

Following the preprocessing and feature-extraction pipeline described above, we evaluated three configurations: (1) a baseline using only transformer-based embeddings with a support-vector machine classifier; (2) a fuzzy-only model applying the Mamdani inference system without generative augmentation; and (3) the full fuzzy–generative framework incorporating synthetic examples from the conditional VAE. All models were trained on 70 % of the data and tested on the remaining 30 %, with five random splits to control for sampling variability.

The fuzzy–generative approach achieved a mean classification accuracy of 84.3 % for critical thinking and 81.7 % for collaboration, outperforming both the fuzzy-only model (75.1 % and 72.4 %, respectively) and the embedding-only SVM baseline (68.9 % and 66.5 %). Cohen's $\kappa$ also improved markedly, rising from 0.56 (baseline) and 0.68 (fuzzy-only) to 0.77 under the full framework, indicating substantial agreement with expert judgments. An ablation study further demonstrated that removing the VAE component led to a 9.2 pp drop in average accuracy, confirming the value of generative augmentation under low-resource conditions.

Error analysis revealed that the combined model was particularly effective at distinguishing medium and high competency levels, where embedding-only classifiers often conflated subtle linguistic cues. Examples of misclassifications primarily occurred in brief messages lacking explicit argumentative structure, suggesting future work to integrate discourse-level features. Overall, the results validate the proposed methodology's ability to provide flexible, transparent, and accurate assessments of implicit competencies from student chat interactions.

The experimental evaluation demonstrates that integrating fuzzy inference with generative augmentation substantially enhances the automated assessment of implicit competencies. The full hybrid framework achieved an average accuracy of 84.3 % for critical thinking and 81.7 % for collaboration, reflecting a marked improvement over both the embedding-only SVM baseline (68.9 % and 66.5 %) and the fuzzy-only variant (75.1 % and 72.4 %). The increase in Cohen's $\kappa$ to 0.77 further indicates strong concordance with expert ratings, confirming that the fuzzy rules effectively translate numerical text features into human-interpretable competency levels and that the conditional VAE successfully alleviates data sparsity where certain competency profiles were underrepresented. In particular, the ablation study revealed a 9.2 pp drop in accuracy without synthetic examples, underscoring the value of generative data augmentation in low-resource educational contexts.

Error analysis suggests that the hybrid model excels at discriminating between medium and high competency tiers by leveraging nuanced sentiment and aspect-term indicators, whereas brief or off-topic messages still pose challenges. These misclassifications point to opportunities for incorporating discourse-level features — such as coherence structure or reply-network metrics — to further refine inference in cases where individual message content is insufficient.

Collectively, these results validate the proposed methodology as a flexible, transparent, and scalable approach for assessing soft skills directly within LMS environments. By combining state-of-the-art NLP embeddings, expert-driven fuzzy logic, and generative model augmentation, the framework not only matches human judgment with high fidelity but also offers a principled path toward continuous improvement and broader applicability across disciplines.

## References

1. Bird S. Natural Language Processing with Python. 1st. ed., O'Reilly, 2009. P. 504.

2. Richardson J.T.E. Instruments for obtaining student feedback: a review of the literature // Assessment & Evaluation in Higher Education. — 2005. — Vol. 30, No. 4. — pp. 387—415.

3. Shaik T. Review of Trends and Challenges in Adopting NLP Methods for Education Feedback Analysis // IEEE Access — 2022. — Vol. 10. — pp. 56720—56739.

4. Gandhi H. Extracting aspect terms using CRF and bi-LSTM models // Procedia Computer Science. — 2020. — Vol. 167. pp. 2486–2495.

5. H. Melnyk, V.Melnyk. Sentiment Analysis for Student Feedback: proceedings of the 2nd International Conference on Innovative Solutions in Software Engineering (ICISSE), Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, Ukraine, 29–30 Nov. 2023. — pp. 99–100.